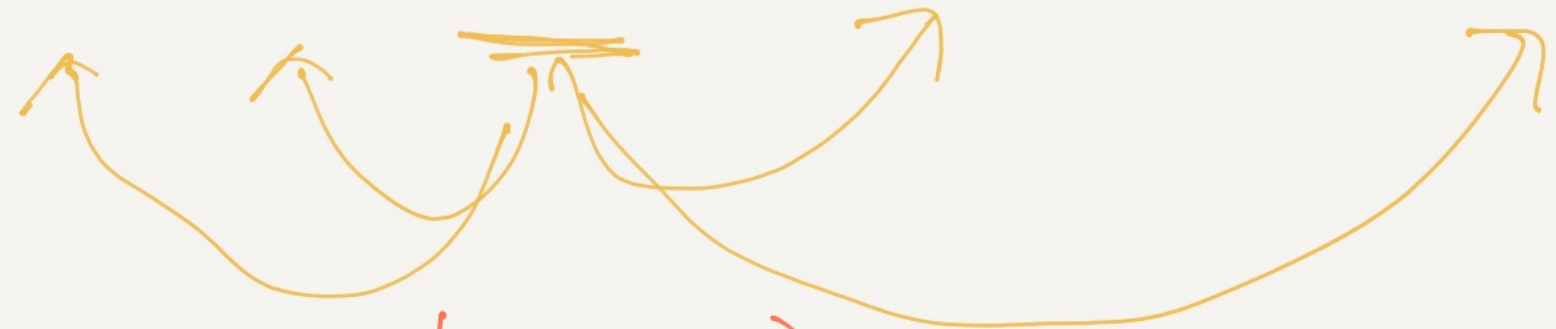2017°3 22   Skip gram in detail.

Last time about skip gram

Udacity is the best education company.

(best, is)

(best, the)

(best, education)

(best, company)

$$P(W_{is} | W_{best})$$

$$P(W_{the} | W_{best})$$

what is target?

$$T' = \prod_{c=1}^{V} \prod_{-m \le j \le m, j \ne 0} P(w_{c+j} \mid w_c)$$

$$T = -\log T' = -\sum_{c=1}^{V} \sum_{\substack{-m \le j \le m \\ j \ne 0}} \log P(w_{c+j} \mid w_c)$$
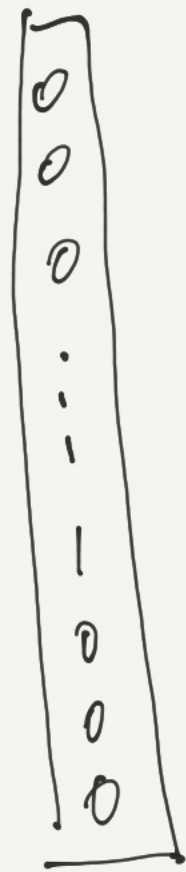
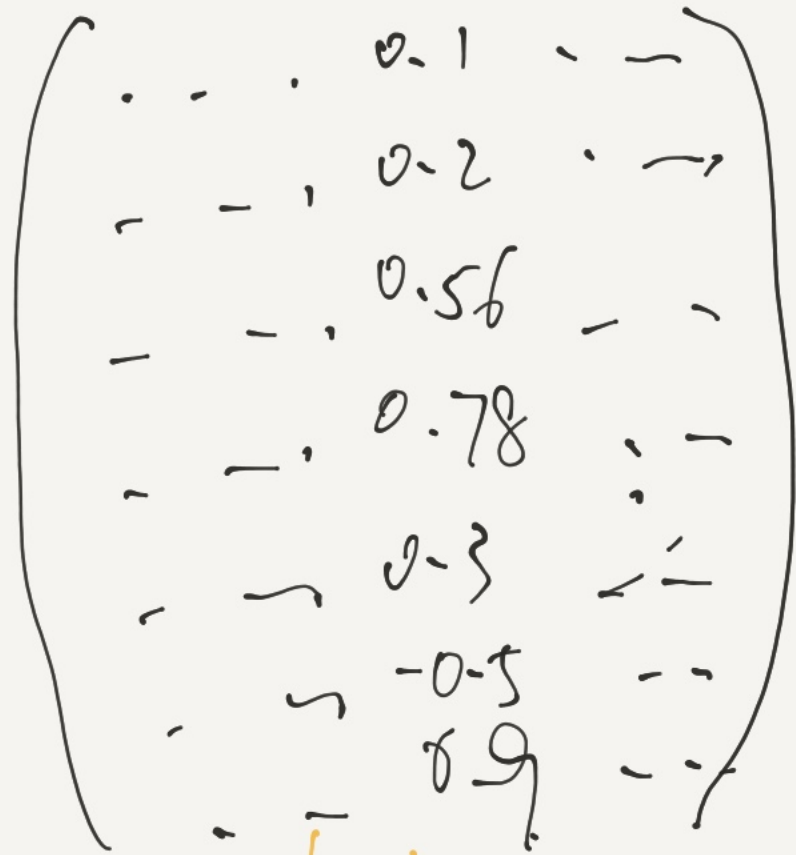$$P(w_o \mid w_I) = \frac{\exp(U_o^T v_I)}{\sum_{w=1}^{V} \exp(U_w^T v_I)}$$

How to compute

???

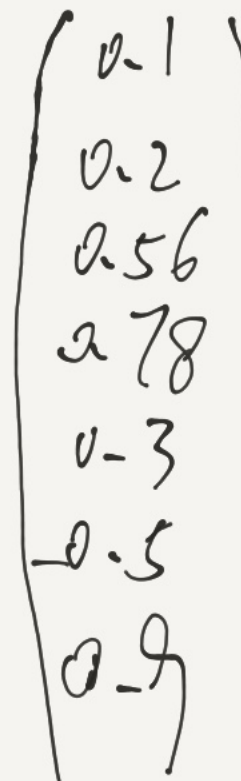# Neural Network to compute target function
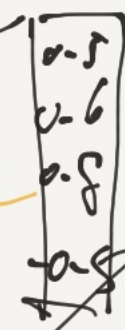

$V \times 1$ $d \times V$ $d \times 1$ $V \times d$ $V \times 1$ cross entropy

softmax

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

one hot encoding

lookup table

$$\begin{bmatrix} \cdots & 0.1 & \cdots \\ \cdots & 0.2 & \cdots \\ \cdots & 0.56 & \cdots \\ \cdots & 0.78 & \cdots \\ \cdots & 0.3 & \cdots \\ \cdots & -0.5 & \cdots \\ & 0.9 & \cdots \end{bmatrix}$$

$$\begin{bmatrix} 0.1 \\ 0.2 \\ 0.56 \\ 0.78 \\ 0.3 \\ -0.5 \\ 0.9 \end{bmatrix}$$

word embedding V_i

u_2

u_3

$$\begin{bmatrix} 0.5 \\ 0.6 \\ 0.8 \\ 0.9 \\ \vdots \end{bmatrix}$$

$$\begin{bmatrix} 0.3 \\ 0.3 \\ 0.35 \\ 0.05 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$
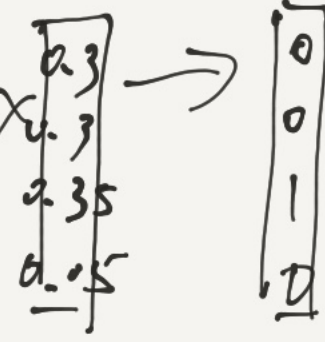
logits   softmax   1 hot encoding

Target function

= cross-entropy (softmax (logit))

How to make it more efficient ?
compute loss function !

Paper Review

1. Sampled softmax

2. Negative Sampling

Sampled softmax loss function

$$\text{loss/target} = -\sum_{c=1}^{V} \sum_{\substack{-m \leq j \leq m \\ j \neq 0}} \log p(w_{c+j} \mid w_c)$$

$$p(w_0 \mid w_I) = \frac{\exp(U_0^T V_I)}{\sum_{w=1}^{V'} \exp(U_w^T V_I)}$$

$$= \frac{\exp(U_0^T V_I)}{\sum_{w=1}^{V'} \exp(U_w^T V_I)} \qquad w_0 \in V'$$

Negative Sampling

$$\log \overset{\times}{P}(w_0 \mid w_I) = \log \sigma(U_0^T V_I) + \sum_{k=1}^{K} \log \sigma(-U_k^T V_I)$$

$$\log P(w_0 \mid w_I) = \log \frac{\exp(U_0^T V_I)}{\sum_{w=1}^{=V} \exp(U_w^T V_I)}$$

maximize.   minimize

$$\sigma(-x) = 1 - \sigma(x)$$